467
SCIENTIFIC AMERICAN OFFPRINTS

ATTENTION AND THE PERCEPTION OF SPEECH

by Donald E. Broadbent

SCIENTIFIC AMERICAN
APRIL 1962
VOL. 206, NO. 4 PP. 143-151

PUBLISHED BY **W. H. FREEMAN AND COMPANY** 660 MARKET STREET, SAN FRANCISCO, CALIFORNIA 94104

Copyright © 1962 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this offprint may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher.

ATTENTION AND THE PERCEPTION OF SPEECH

If an individual listens to two voice messages at one time, he usually understands only one. This indicates that the brain has an "attention mechanism" for selecting the desired information

by Donald E. Broadbent

Paying attention—and not paying attention—are surely two of the most important abilities of human beings. Yet in spite of their crucial role in learning, and in a host of other intelligent activities, psychologists for many years did not consider them proper topics of study. Attention seemed a subjective quality, associated historically with the introspective method of investigation. That method tends to give inconsistent results and so fell into disrepute among experimental psychologists. Correspondingly, most respectable theorists failed to make use of any concept resembling attention; and, since research in psychology tends to be dominated by theory, there was little experimentation along lines that might have revived the idea.

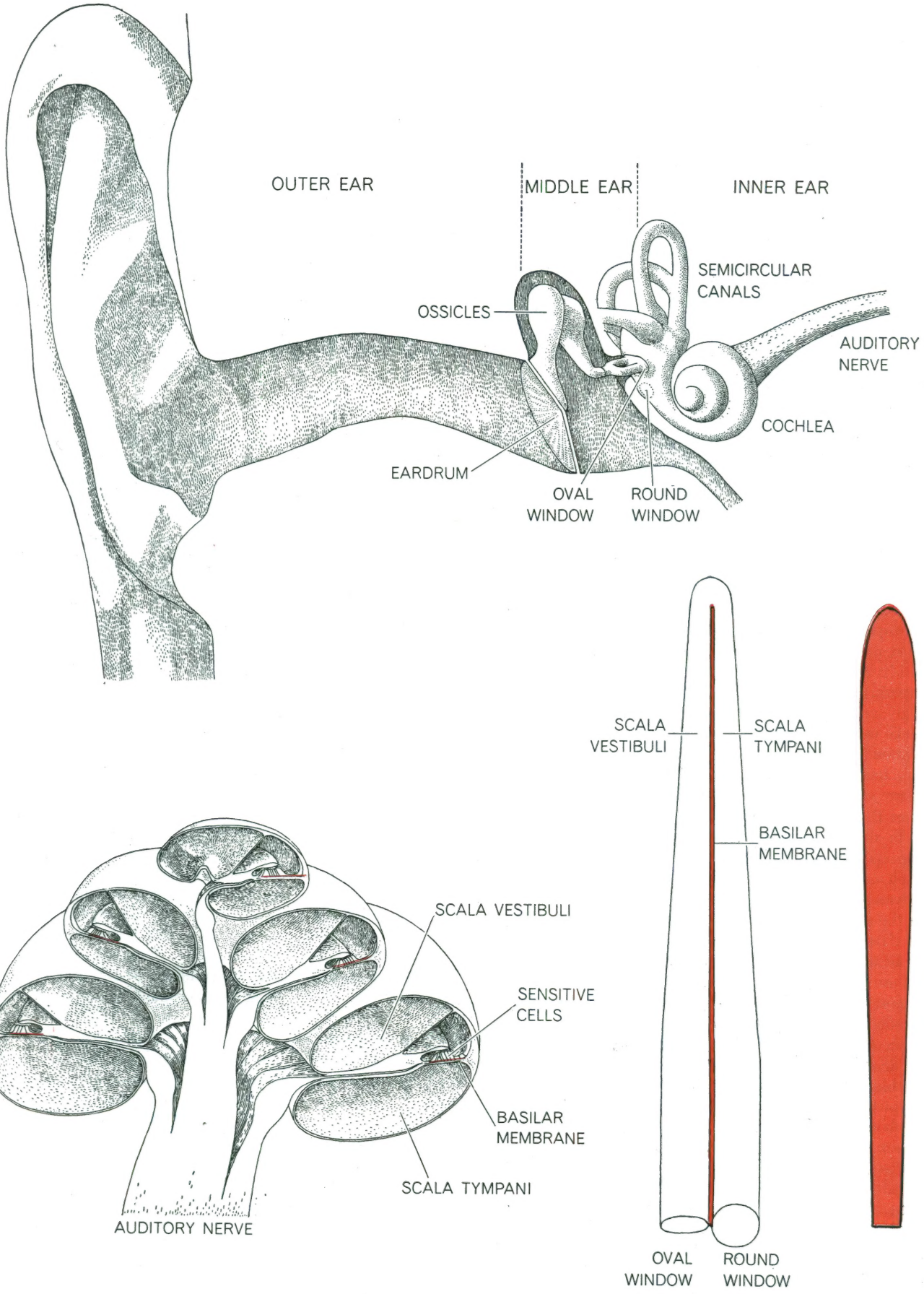
In the past 10 years, however, the concept of attention has begun to force itself on the attention of psychologists in various ways. One is through studies of the efficiency of control systems such as those concerned with the regulation of air traffic at airports. A major cause of failure in these systems is that the human operator has too much information to handle simultaneously, or that he reacts to an unimportant signal when he should be dealing with an important one. These problems require some understanding of phenomena that would commonly be described under the heading of "attention." There is now accumulating a wide variety of experimental results that clarify these phenomena, although the larger part of the work remains to be done. In this article I shall describe some of the research on attention to spoken messages.

One of the earliest findings, and one that agrees with everyday experience, is that it is harder to understand two messages arriving simultaneously than two messages arriving one after the other. One might be tempted to explain this as a purely physical interference between the two stimuli; for example, the louder passages of one message might drown out the softer passages of the other and vice versa, rendering them both unintelligible. Actually the matter is not so simple. By recording the messages on tape and playing them for different subjects instructed to respond in different ways, the intelligibility is shown to depend on psychological factors. Specifically, either message becomes understandable if the listener is instructed to ignore the other. But the two messages together cannot both be understood, even though the necessary information is available to the ear. Another way of making the same point is to insert the words of one message into spaces between the words from the other: "Oh God say save can our you gracious see Queen." Each message is hard to understand, but each word is spoken separately and is fully audible. The difficulty evidently lies inside the nervous system, which somehow prevents an adequate response to signals that are "heard" satisfactorily.

Further experiments demonstrate that comprehension improves if the two messages differ in certain physical characteristics. For instance, it is better if a man speaks one message and a woman speaks the other; or if the loudspeaker removes the lower tones from one voice but not the other. Spatial separation of the two voices gives the best result of all. The different messages should not come through the same loudspeaker or even from separate speakers mounted one above the other; the two speakers should be separated as far as possible from each other in the horizontal plane. Interestingly enough, a listener also comprehends simultaneous spoken messages better when they come from a stereophonic system than when they are played over a single loudspeaker. (This effect, rather than the doubtful gain in realism, is for many people the main advantage of stereophonic high-fidelity systems: the listener can pay attention to different musical instruments played at the same time.)

Physical distinctions are most helpful in promoting understanding when one message has no importance for the listener and does not have to be answered. It would seem that the differences allow the brain to filter the incoming sounds and select some for response while ignoring others.

The need to throw away part of the available information can perhaps be understood by comparing the brain with man-made communication systems. Engineers nowadays talk of capacity for transmitting information, by which they mean the number of equally probable messages of which one can be sent in a specified time. Suppose, for example, that two complicated military plans have been prepared and an order is to be sent to carry out one of them. A simple communication system consisting of a red and a green lamp can transmit the message with maximum efficiency by the lighting of a single lamp. If there were four plans instead of two, however, it would be impossible to give the order by lighting one of the two lamps no matter how simple each plan might be. Either there must be more lamps or more time is needed for sending the order. In the most efficient code for two lamps, two successive flashes of the red lamp would mean one plan, a red flash followed by a green flash would mean another, and so on. One of four possible messages can be transmitted with two lamps, but only by taking two units of time. With eight pos-



PERCEPTION OF SPEECH begins in the ear, shown at top in simplified cross section. The eardrum transmits sound vibrations to the three small bones called ossicles, which cause waves in fluid in the cochlea. The cochlea, seen in cross section at bottom left, contains the basilar membrane (*color*), on which rest the sensitive cells that excite auditory-nerve fibers. At bottom center cochlea is rolled out, with basilar membrane in side view. Front view of the basilar membrane (***bottom right***) shows that it is wider at one end than the other. The wide region vibrates in response to low frequencies, whereas the narrow region responds to high frequencies.

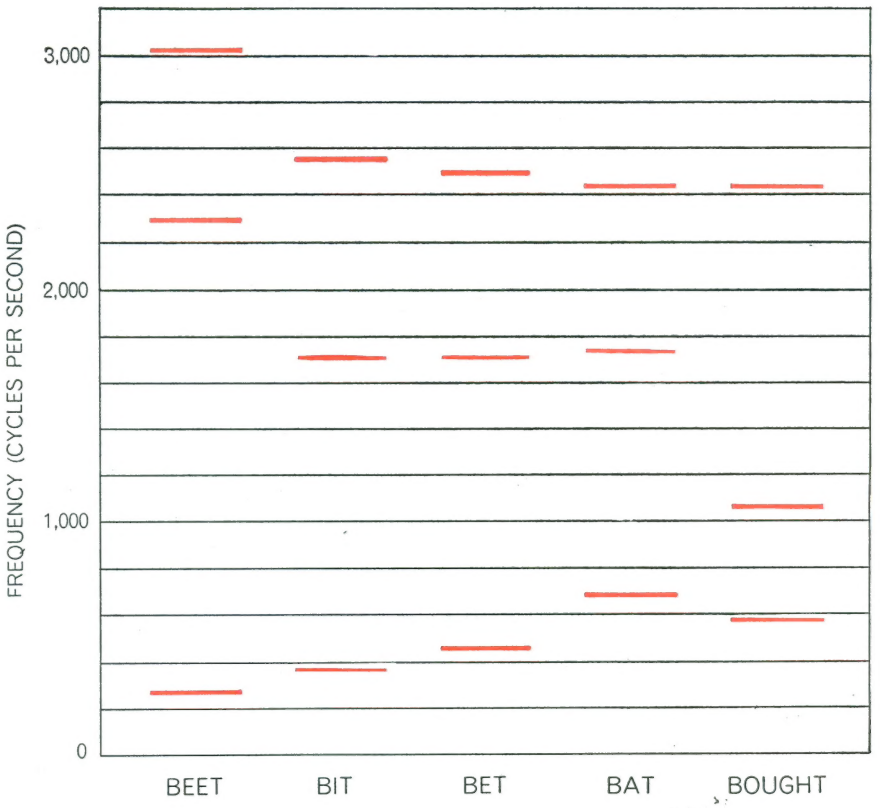
sible messages the code would call for three flashes of the two lamps, taking three units of time; 16 possible messages would require four flashes, and so forth.

Although the human brain has far more than the two states represented by the red and the green lamp, the number of its possible states is presumably limited. One would expect, then, that there is a limit to the number of different possibilities among which it can distinguish in a given time. Indeed, a number of experiments suggest a close parallel with the two-lamp system: in many cases a man's reaction time in responding to one of several possible signals increases by an equal amount every time the number of possible signals is doubled. Since there is a maximum speed at which one signal can be distinguished from others, the brain limits the number of possibilities being considered at any one time by selecting only part of the information reaching the ears. Therefore the degree of difficulty in dealing with two simultaneous spoken messages depends on the number of other messages that might have arrived instead of the two that did arrive. If only a few other messages are possible, the two messages together may not exceed the capacity of the brain and the listener may understand both. On the other hand, if each message is drawn from a very large range of possibilities, it may be all the listener can do to respond appropriately to one of them.

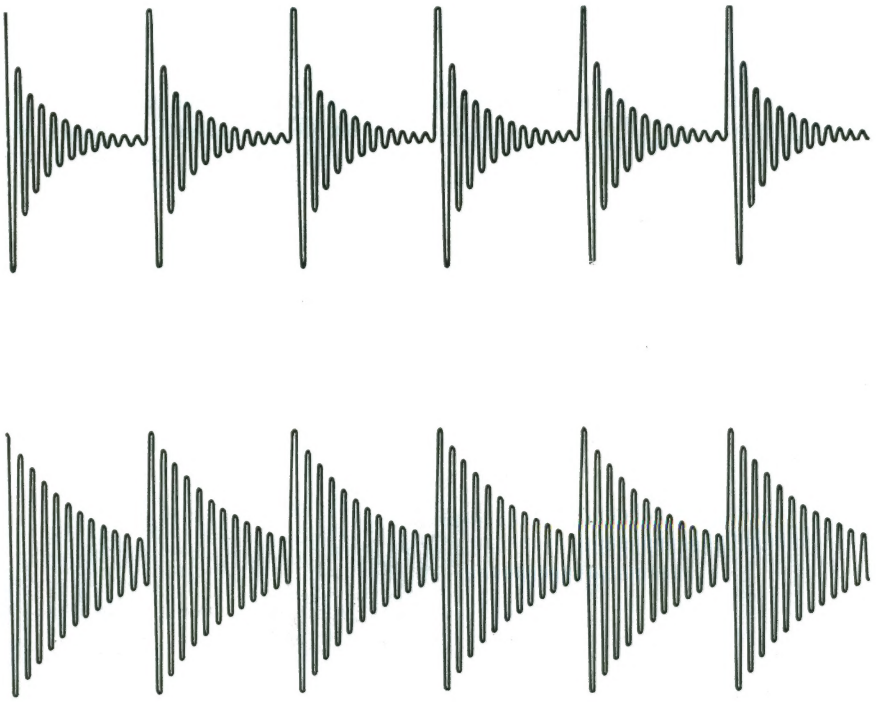
Several studies support these conclusions. John C. Webster and his associates at the U.S. Navy Electronics Laboratory in San Diego, Calif., observed that control-tower operators in San Diego could sometimes identify two aircraft call signs arriving at the same time but could understand only one of the two messages that followed. The call signs penetrated because the operators knew pretty well which aircraft might call. They did not know what the pilots would say.

An experiment at the Applied Psychology Research Unit in Cambridge, England, required a listener to answer a rapid series of questions while pressing a key in response to an intermittent buzzer. The interference produced by the buzzer in the ability to answer questions increased after the subject had been told that he would also have to respond to a bell. Even when the bell did not ring, the subject found the questions harder to answer than when he was expecting only the buzzer.

These results help to explain why a person can sometimes listen to two things at once and sometimes cannot pay atten-



VOWEL FORMANTS, or frequencies that make up each vowel sound, are shown here for five different vowels. The values given are averages for male voices. Actually they differ from person to person. Although three formants are shown here for each sound, quite recognizable vowels can be produced by mechanisms using two filters to make two formants.



DECAY RATE of pulses from vocal cords affects quality of speech. Waves of highest amplitude mark beginning of each pulse. At top, pulses decay rapidly, helping to give the voice a crisp or sharp sound. At bottom, the decay is much slower, giving the voice a mellow quality. In both cases the frequencies of the pulses and the vibrations are exactly the same.

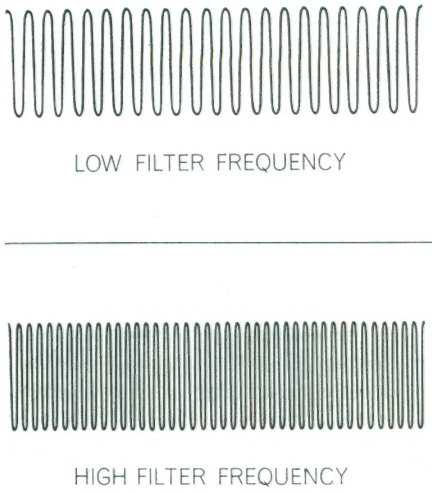
tion to more than one. When the listener is thoroughly familiar with a situation, so that he knows to within a small number of alternatives what each message will be, he can comprehend two simultaneous messages. But when one or both messages are drawn from a large number of possibilities, the filter in the brain lets only one message come through.

How does the filter work? As yet the answer is not known. Enough is known, however, about the physical characteristics of speech and the physiology of hearing to make possible some reasonable speculation. Human speech is produced by the combined action of the vocal cords and the vocal tract, which consists of the cavities of the throat, mouth and nose. Taut vocal cords produce a buzz when air is forced through them. The buzz consists of brief pulses, or puffs of air, at the rate of 100 or more per second, each pulse containing energy at many frequencies. These pulses excite into vibration the air in the cavities of the throat, nose and mouth. The cavities can be tuned to different frequencies by changing the position of the tongue, cheeks, jaw and lips. What emerges is a train of waves that contains a particular group of frequencies and is pulsed about 100 times per second. Each pulse starts out at full strength and decays rapidly until the sound energy is renewed by the next one [*see bottom illustration on preceding page*].

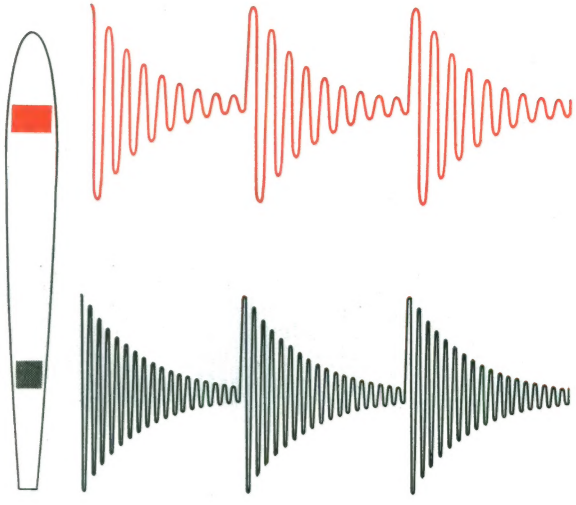
Many vowel sounds contain waves at two or more widely separated frequencies. For example, when the greatest energy is at 375 and 1,700 cycles per second, the vowel sound in the word "bit" is produced; frequencies of 450 and 1,700 cycles per second give the vowel in "bet." (These figures apply to a typical male voice. In the voices of women and children the whole range of frequencies may be higher but the listener takes this into account.) On reaching the ear, the sounds stimulate sense organs arranged along the basilar membrane in the cochlea [*see illustration on page 4*]. Low frequencies stimulate organs at one end of the membrane; high frequencies affect those at the other end. A complex sound made up of several frequencies energizes several different regions of the basilar membrane. Each sense organ on the membrane connects with particular nerve fibers going to the brain; thus the word "bit" stimulates one combination of fibers and the word "bet" another combination.

If both words reach the ear simultaneously, both combinations of fibers would come into play and the brain would have the problem of deciding which belong together. It might seem then that two or more voices would produce so much confusion in the ear that the brain could not select one voice for special attention. Of course, certain obvious features help distinguish one speaker from another: accent, rate of speaking, loudness or softness. But one cannot make use of these features until one knows which frequencies belong to which voice. Thus the problem remains: How does the brain manage to focus attention on one voice? Studies of the artificial generation of speech sounds have begun to throw some light on this problem.

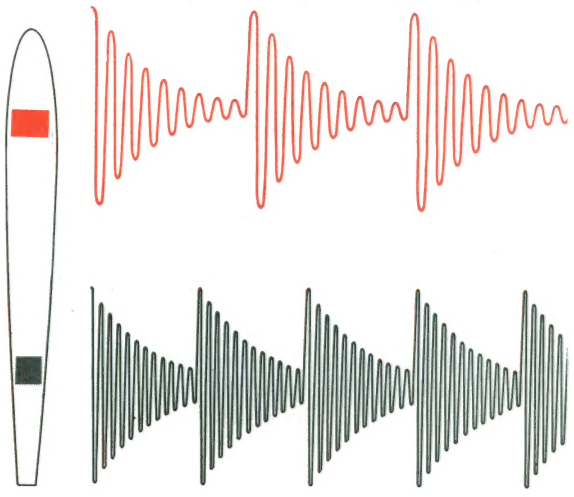
Peter Ladefoged of the University of Edinburgh and I have been experimenting with a device that was developed by Walter Lawrence of the Signals Research Development Establishment in England. Our version of the apparatus sends a series of electrical pulses (analogous to pulses from the vocal cords) through two filter circuits, each of which passes primarily one frequency. The waves from one filter circuit, which are like those from the largest human speech cavity, are mixed with waves from the other, which imitate the frequencies produced by the second largest cavity. To-



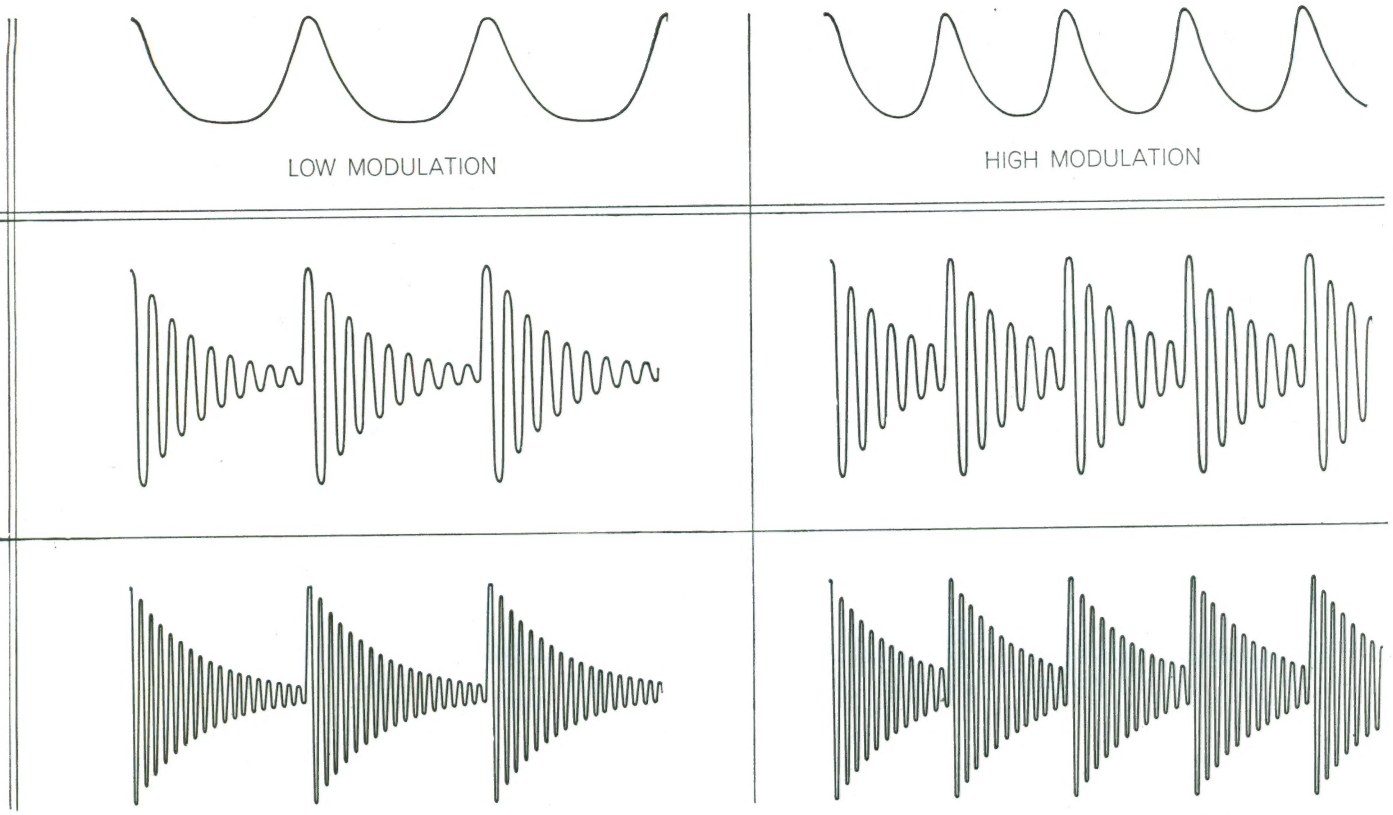
SPEECH SOUNDS consist of pulses of energy from the voice, shown here as high and



DIFFERENT FILTER FREQUENCIES but same rate of pulsation or modulation from voice excites two different regions of a basilar membrane (*left*). Listener reports he hears one vowel sound.



DIFFERENT PULSATION RATES and different filter frequencies make the listener hear two different sounds, even though only one ear or basilar membrane is actually being used for hearing.

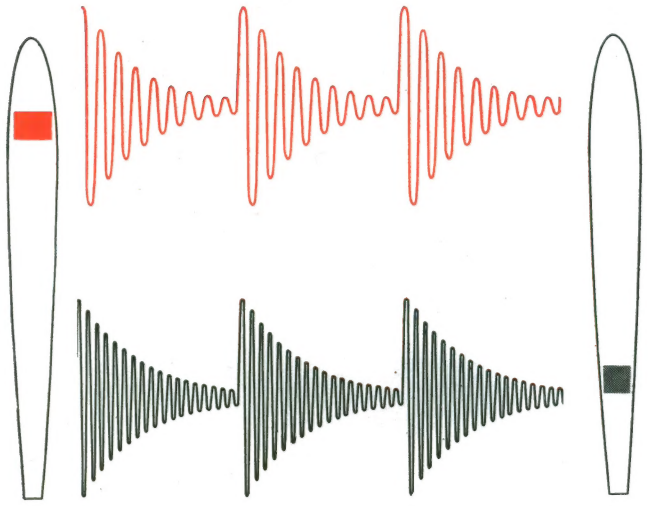


low modulations (*across top*), and of specific frequencies emitted by the mouth and throat "filters," or cavities (*far left*). These two types of wave combine in patterns like those in this diagram. Effects of such waves are shown across bottom of these two pages.

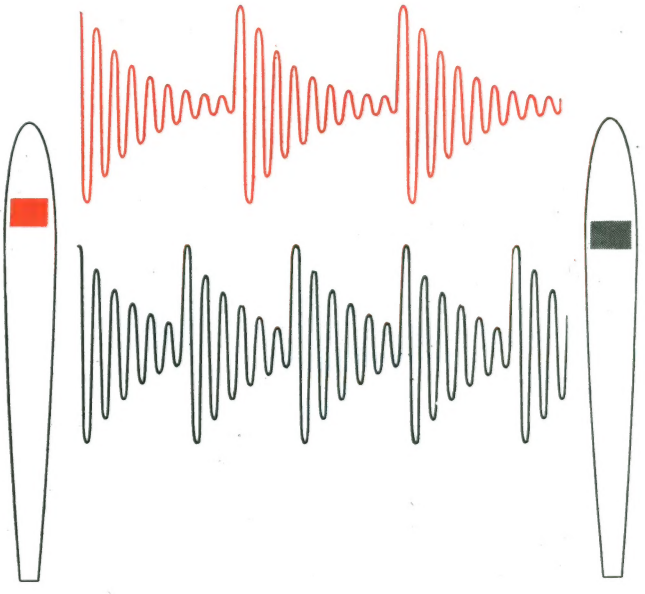
gether the two wave trains are heard as quite acceptable vowel sounds that can be changed by tuning the filters to different frequencies. Varying the pulse rate used to excite the filters alters the apparent pitch or intonation of the "speech": it rises with faster pulse rates and falls with slower ones.

When the same pulses excite both filters, a listener hears the output as readily identifiable vowel sounds. This is true even when the low frequency is fed into one ear and the high frequency into the other. But if the two filters are pulsed at slightly different rates, the "speech" becomes unacceptable and listeners say that they are hearing two sounds coming from two sources rather than a single vowel sound.

Other experiments on the fusion of sounds at the two ears, conducted by Colin A. Cherry and his colleagues at the Imperial College of Science and Technology in London, also support the idea that when the rate of pulsing, or modulation, is the same for two sounds, the hearer perceives them as one sound. It seems reasonable to suppose, therefore, that a man can listen to one person



USING BOTH EARS, listener will hear one vowel sound, although right ear hears one filter frequency and left ear hears another. The pulsation or modulation rate has to be the same.

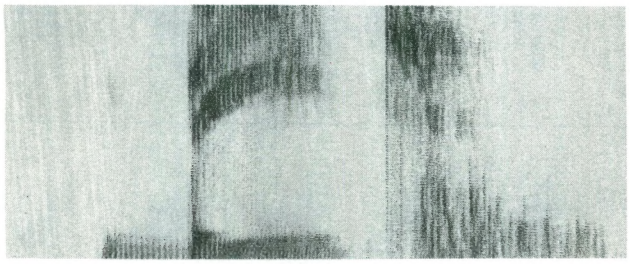


TWO PULSATION RATES, combined with same filter frequency and fed into each ear separately, produce two distinct sounds. The brain evidently focuses its attention on the rate of pulsation.

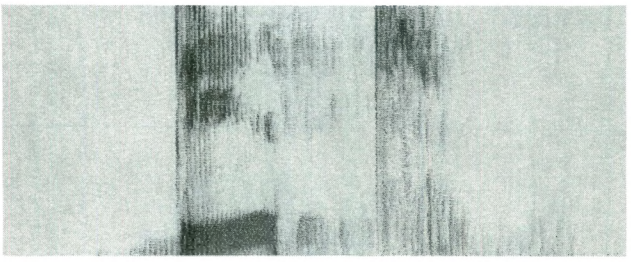
and ignore another primarily by selecting from the mass of sounds entering his ears all those frequencies that are being modulated at the same rate. Since it is most unlikely that the vocal cords of two speakers would vibrate at exactly the same rate at any moment, modulation would almost always provide an important (if not the sole) means of separating a pair of voices.

It is now a generally accepted principle of neurophysiology that messages traveling along a particular nerve can differ either by involving different nerve fibers or by producing a different number of impulses per second in the fibers. High-frequency and low-frequency sounds stimulate different fibers. It may be that the rate at which the sounds are pulsed controls the rate of firing of the fibers. If so, the brain could pick out one voice from others by focusing its attention on all auditory nerve fibers that are firing at the same rate.

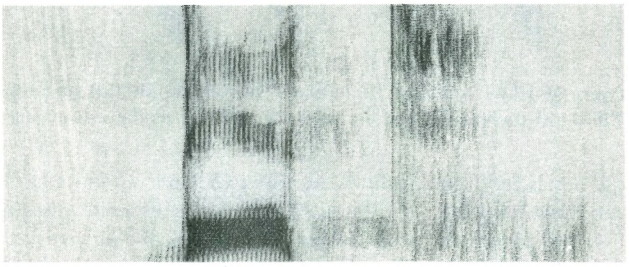
A further indication of the importance of modulation is that it, rather than the frequency of the waves being modulated, seems under certain conditions to determine the pitch of a voice. This can be demonstrated with the artificial speech generator. A filter tuned to, say, 3,000 cycles per second is pulsed at the rate of 100 cycles per second. A listener is asked to match the pitch of the sound with either of two simple sound waves, one at 100 cycles per second and the oth-



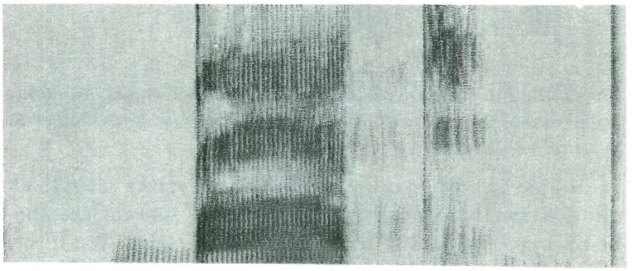
"BEET"



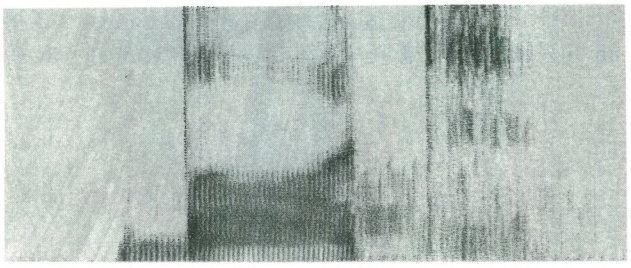
"BIT"



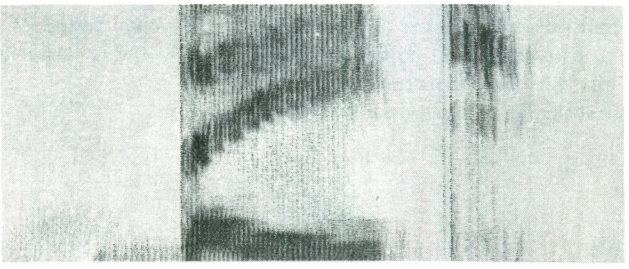
"BET"



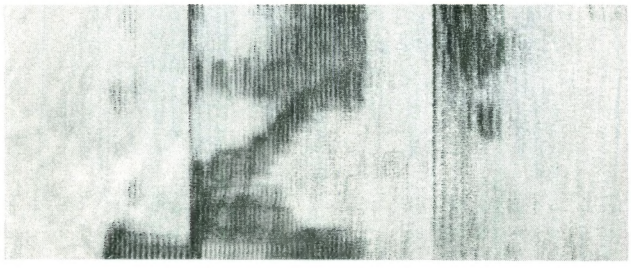
"BAT"



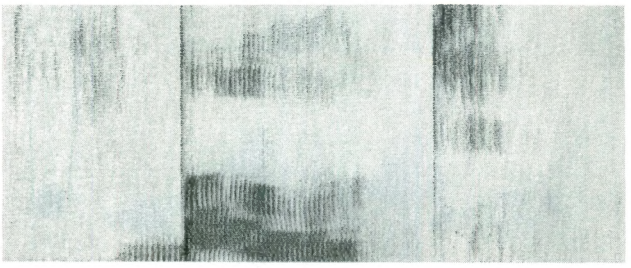
"BOUGHT"



"BATE"

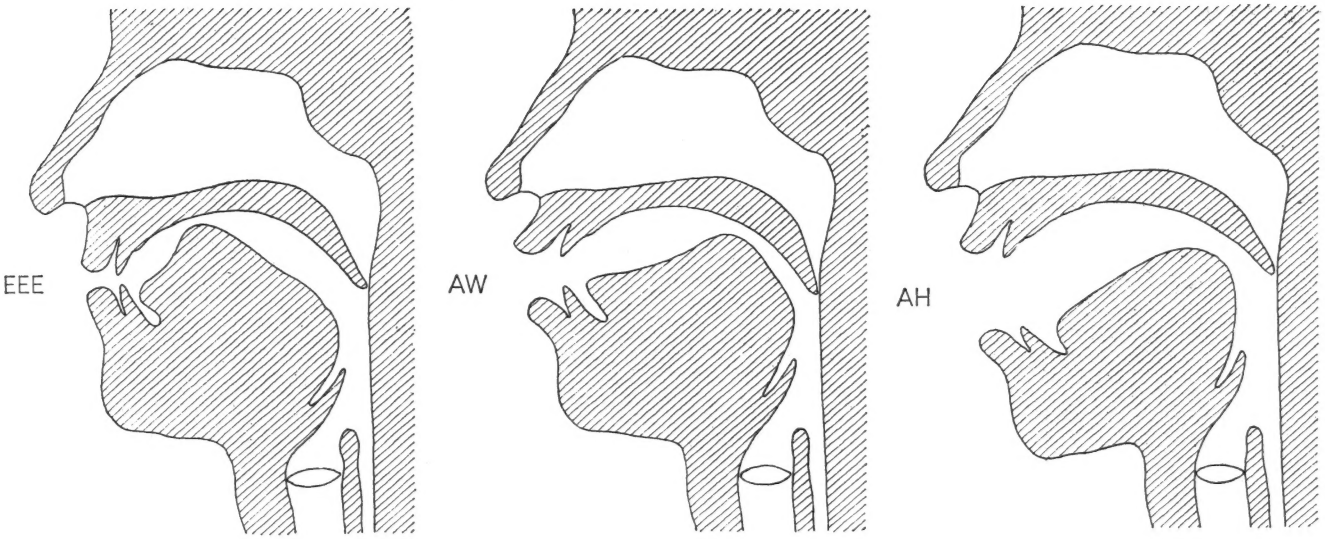


"BITE"



"BOAT"

SOUND SPECTROGRAMS show that various vowel sounds are made of several different frequencies. Time is shown horizontally, frequencies vertically and intensity of sound by relative darkness. The "b" of each word appears at lowest frequency. Vowel begins suddenly as lips open. After vowel there is a quiet period followed by a burst of noise primarily at high frequency as the "t" explodes. Frequency shifts in "bate" and "bite" are diphthongs. Spectrograms were made by H. K. Dunn of Bell Telephone Laboratories.



SHAPE OF CAVITIES in the mouth is primarily responsible for the production of different vowel sounds. Three other factors that play a key role in this process are the configuration of the tongue, the size of the opening of the mouth and the position of the lips.

er at 3,000. Usually he selects the 100-cycle sound.

The selection mechanism that has been described is still hypothetical, but I believe that something much like it must exist. There can be no doubt, however, that it is not the only basis for auditory attention. Several experiments have served to make this clear. In one, a listener is equipped with earphones that feed one voice into the right ear and another voice into the left. Normally the subject has no difficulty in understanding the message entering one ear and ignoring the other. But under certain conditions sound from the ear being ignored can break into consciousness. For example, Neville Moray of the University of Oxford has demonstrated that a man fully occupied in listening to speech entering one ear will hear his own name in the other ear even though he remains quite unresponsive to any other word in that ear. Under similar circumstances Anne M. Treisman of the University of Oxford has found that speech entering the rejected ear can break through to the subject's attention if it consists of words that would probably follow the words that have just been heard by the ear that is receiving attention. In these cases the content of the speech has taken precedence over its physical characteristics.

How the brain focuses attention on meaning or content is as yet an almost complete mystery. One thing is clear. If the method proposed for choosing between voices is correct, there must be two attention mechanisms. Selection on the basis of content involves examining a stimulus for its possible appropriateness to a particular set of responses rather than for the presence or absence of a physical marker. At one moment, for example, a person might be ready to write down any of the digits one through nine and highly unready to write anything else, or indeed to respond in any other way. If he hears a sound from any direction or in any voice that can be interpreted as the name of one of the digits, he will respond by writing it down; only if the sound cannot be so interpreted will he not respond. At another time he might be ready to write down letters of the alphabet but not numbers, and so on.

Both types of attention are now the subject of intensive research. The next few years should yield more definite clues to the nature of each and at least a tentative answer to the question of whether or not they depend on different mechanisms.

The Author

DONALD E. BROADBENT is director of the Applied Psychology Research Unit of the Medical Research Council in Cambridge, England. His original intention was to be a physical scientist, but while serving with the Royal Air Force he became interested in the problem of how the arrangement of controls on instrument panels might make some aircraft more difficult to fly than others. Subsequently he took a degree in psychology at the University of Cambridge in 1949, going to work in the Applied Psychology Research Unit immediately thereafter. Broadbent became director of the Applied Psychology Research Unit in 1958.

Bibliography

CONTEXTUAL CUES AND SELECTIVE LISTENING. Anne M. Treisman in *Quarterly Journal of Experimental Psychology,* Vol. 12, No. 4, pages 242–248; November, 1960.

ON THE FUSION OF SOUNDS REACHING DIFFERENT SENSE ORGANS. D. E. Broadbent and Peter Ladefoged in *The Journal of the Acoustical Society of America,* Vol. 29, No. 6, pages 708–710; June, 1957.

PERCEPTION AND COMMUNICATION. D. E. Broadbent. Pergamon Press, Inc., 1958.

THREE AUDITORY THEORIES. J. C. R. Licklider in *Psychology: A Study of a Science,* edited by Sigmund Koch, Vol. 1, pages 41–144. McGraw-Hill Book Company, Inc., 1959.

Study Guide

Prepared by JOHN P. J. PINEL, UNIVERSITY OF BRITISH COLUMBIA

ATTENTION AND THE PERCEPTION OF SPEECH

Donald E. Broadbent APRIL 1962

SCIENTIFIC AMERICAN OFFPRINT 467

I. SUMMARY

We only respond to a small proportion of stimuli which affect our sensory receptors. The study of this sensory selectivity is commonly subsumed under the heading of attention.

It is more difficult to understand two verbal messages if they are presented simultaneously than if they are presented sequentially, and this is not because of direct physical interference between the stimuli, since either message becomes perfectly intelligible if the subject is instructed to selectively attend to it. This so-called selective attention is facilitated if the physical characteristics of the two messages differ. For example, one message has little effect on the comprehension of the concomitant message if they differ markedly in terms of pitch or position of source. Selective attention is also easier when the listener is thoroughly familiar with the situation. In certain situations where one knows that messages are selected from only a small number of alternatives, it is possible to comprehend two simultaneous messages. However, when one or both alternatives are drawn from a large number of possibilities, the brain processes only one.

Speech is composed of complex pulses of sound, each pulse containing energy at many frequencies. When two or more messages are presented simultaneously, on what basis does the brain filter out all the sounds except those associated with a single message? The results of studies employing the artificial generation of speech sounds indicate that words with similar patterns of pulsing are perceived as belonging to the same message. Thus, the brain seems to filter certain messages on the basis of pulsation pattern regardless of sound frequency. Although this filtering mechanism is remarkably effective, there are certain stimuli such as the listener's name which always seem to by pass the filter.

II. GLOSSARY

auditory nerve – a bundle of neurons transmitting coded auditory inputs to the brain.
basilar membrane – membrane which houses the auditory receptors and divides the cochlea into scala vestibuli and scala tympani.
cochlea – spiral-shaped organ of the inner ear containing the basilar membrane, scala vestibuli, and scala tympani.
introspection – attempting to objectively observe the activities of one's own mind.
ossicles – bones of the middle ear which transmit vibrations from the eardrum to the oval window.
oval window – transmits vibrations from the ossicles to the fluids of the cochlea.
round window – absorbs hydraulically the pressures produced in the cochlea.
scala tympani – the half of the cochlea containing the round window.
scala vestibuli – the half of the cochlea containing the oval window.

III. ESSAY STUDY QUESTIONS

1. Explain one of the ways the concept of "attention" has forced itself on experimental psychologists.
2. The fact that two simultaneous verbal messages can not both be comprehended is not attributable to a physical interference between the two stimuli. Explain.
3. Describe the physical characteristics of speech.
4. Describe the experiments by Broadbent and Ladefoged which indicate that filtering is based on the rate of pulsation.
5. Some words are filtered out; others are not. In order to determine which words are to be filtered, the brain must receive information about the words. Thus, paradoxically the brain must "hear" a word before it can filter it out so that it will not be heard. Discuss.
6. What are the two kinds of attention mechanisms?